

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Bradfield, Christopher A.
Dolwick, Kristin M.
Carver, Lucy A.
- (ii) TITLE OF INVENTION: Ah Receptor cDNAs and Genetically Engineered Cells for Detecting Agonists to the Ah Receptor
- (iii) NUMBER OF SEQUENCES: 36
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut
(B) STREET: 100 South Wacker Drive, Suite 960
(C) CITY: Chicago
(D) STATE: Illinois
(E) COUNTRY: USA
(F) ZIP: 60606-4002
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Tilton, Timothy L.
(B) REGISTRATION NUMBER: 16,926
(C) REFERENCE/DOCKET NUMBER: NU-9207-CIP
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (312)-456-8000
(B) TELEFAX: (312)-456-7776

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3207 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

0885402.051397

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..2415

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG AGC AGC GGC GCC AAC ATC ACC TAT GCC AGC CGC AAG CGG CGC AAG
48
Met Ser Ser Gly Ala Asn Ile Thr Tyr Ala Ser Arg Lys Arg Arg Lys
1 5 10 15
CCG GTG CAG AAA ACA GTA AAG CCC ATC CCC GCT GAA GGA ATT AAG TCA
96
Pro Val Gln Lys Thr Val Lys Pro Ile Pro Ala Glu Gly Ile Lys Ser
20 25 30
AAT CCT TCT AAG CGA CAC AGA GAC CGG CTG AAC ACA GAG TTA GAC CGC
144
Asn Pro Ser Lys Arg His Arg Asp Arg Leu Asn Thr Glu Leu Asp Arg
35 40 45
CTG GCC AGC CTG CTG CCC TTC CCG CAA GAT GTT ATT AAT AAG CTG GAC
192
Leu Ala Ser Leu Leu Pro Phe Pro Gln Asp Val Ile Asn Lys Leu Asp
50 55 60
AAA CTC TCT GTT CTT AGG CTC AGC GTC ACG TAC CTG AGG GCC AAG AGC
240
Lys Leu Ser Val Leu Arg Leu Ser Val Thr Tyr Leu Arg Ala Lys Ser
65 70 75 80
TTC TTT GAT GTT GCA TTA AAG TCC ACC CCT GCT GAC AGA AAT GGA GGC
288
Phe Phe Asp Val Ala Leu Lys Ser Thr Pro Ala Asp Arg Asn Gly Gly
85 90 95
CAG GAC CAG TGT AGA GCA CAA ATC AGA GAC TGG CAG GAT TTG CAA GAA
336
Gln Asp Gln Cys Arg Ala Gln Ile Arg Asp Trp Gln Asp Leu Gln Glu
100 105 110
GGA GAG TTC TTG TTA CAG GCG CTG AAT GGC TTT GTG CTG GTT GTC ACA
384
Gly Glu Phe Leu Leu Gln Ala Leu Asn Gly Phe Val Leu Val Val Thr
115 120 125
GCA GAT GCC TTG GTC TTC TAT GCT TCC TCC ACT ATC CAA GAT TAC CTG
432

265550-2015580

Ala 130	Asp	Ala	Leu	Val	Phe	Tyr 135	Ala	Ser	Ser	Thr	Ile 140	Gln	Asp	Tyr	Leu
GGC 480	TTT	CAG	CAG	TCT	GAT	GTC	ATC	CAT	CAG	AGC	GTA	TAT	GAG	CTC	ATC
Gly 145	Phe	Gln	Gln	Ser	Asp 150	Val	Ile	His	Gln	Ser 155	Val	Tyr	Glu	Leu	Ile 160
CAT 528	ACA	GAA	GAC	CGG	GCG	GAA	TTC	CAG	CGC	CAG	CTT	CAC	TGG	GCT	CTA
His	Thr	Glu	Asp	Arg 165	Ala	Glu	Phe	Gln	Arg 170	Gln	Leu	His	Trp	Ala 175	Leu
AAC 576	CCA	GAC	TCT	GCA	CAA	GGA	GTG	GAC	GAA	GCC	CAT	GGC	CCT	CCA	CAG
Asn	Pro	Asp	Ser 180	Ala	Gln	Gly	Val	Asp 185	Glu	Ala	His	Gly	Pro 190	Pro	Gln
GCA 624	GCA	GTC	TAT	TAT	ACC	CCA	GAC	CAG	CTT	CCT	CCA	GAG	AAC	GCT	TCT
Ala	Ala	Val 195	Tyr	Tyr	Thr	Pro	Asp 200	Gln	Leu	Pro	Pro 205	Glu	Asn	Ala	Ser
TTC 672	ATG	GAG	AGG	TGC	TTC	AGG	TGC	CGG	CTG	AGG	TGC	CTG	CTG	GAT	AAT
Phe 210	Met	Glu	Arg	Cys	Phe	Arg 215	Cys	Arg	Leu	Arg	Cys 220	Leu	Leu	Asp	Asn
TCA 720	TCT	GGT	TTT	CTG	GCA	ATG	AAT	TTC	CAA	GGG	AGG	TTA	AAG	TAT	CTT
Ser 225	Ser	Gly	Phe	Leu	Ala 230	Met	Asn	Phe	Gln	Gly 235	Arg	Leu	Lys	Tyr	Leu 240
CAT 768	GGA	CAG	AAC	AAG	AAA	GGG	AAG	GAC	GGA	GCG	CTG	CTT	CCT	CCA	CAA
His	Gly	Gln	Asn	Lys 245	Lys	Gly	Lys	Asp	Gly 250	Ala	Leu	Leu	Pro	Pro 255	Gln
CTG 816	GCT	TTG	TTT	GCA	ATA	GCT	ACT	CCA	CTT	CAG	CCA	CCC	TCC	ATC	CTG
Leu	Ala	Leu	Phe 260	Ala	Ile	Ala	Thr 265	Pro	Leu	Gln	Pro	Pro	Ser 270	Ile	Leu
GAA 864	ATT	CGA	ACC	AAA	AAC	TTC	ATC	TTC	AGG	ACC	AAA	CAC	AAG	CTA	GAC
Glu	Ile	Arg 275	Thr	Lys	Asn	Phe	Ile 280	Phe	Arg	Thr	Lys	His 285	Lys	Leu	Asp
TTC 912	ACA	CCT	ATT	GGT	TGT	GAT	GCC	AAA	GGG	CAG	CTT	ATT	CTG	GGC	TAT
Phe	Thr 290	Pro	Ile	Gly	Cys	Asp 295	Ala	Lys	Gly	Gln	Leu 300	Ile	Leu	Gly	Tyr

ACA GAA GTA GAG CTG TGC ACA AGA GGA TCG GGG TAC CAG TTC ATC CAT
 960
 Thr Glu Val Glu Leu Cys Thr Arg Gly Ser Gly Tyr Gln Phe Ile His
 305 310 315 320

GCT GCA GAC ATA CTT CAC TGT GCA GAA TCC CAC ATC CGC ATG ATT AAG
 1008
 Ala Ala Asp Ile Leu His Cys Ala Glu Ser His Ile Arg Met Ile Lys
 325 330 335

ACT GGA GAA AGT GGC ATG ACA GTT TTC CGG CTT CTT GCA AAA CAC AGT
 1056
 Thr Gly Glu Ser Gly Met Thr Val Phe Arg Leu Leu Ala Lys His Ser
 340 345 350

CGC TGG AGG TGG GTC CAG TCC AAT GCA CGC TTG ATT TAC AGA AAT GGA
 1104
 Arg Trp Arg Trp Val Gln Ser Asn Ala Arg Leu Ile Tyr Arg Asn Gly
 355 360 365

AGA CCA GAT TAC ATC ATC GCC ACT CAG AGA CCA CTG ACG GAT GAA GAA
 1152
 Arg Pro Asp Tyr Ile Ile Ala Thr Gln Arg Pro Leu Thr Asp Glu Glu
 370 375 380

GGA CGA GAG CAT TTA CAG AAG CGA AGT ACG TCG CTG CCC TTC ATG TTT
 1200
 Gly Arg Glu His Leu Gln Lys Arg Ser Thr Ser Leu Pro Phe Met Phe
 385 390 395 400

GCT ACC GGA GAG GCT GTG TTG TAC GAG ATC TCC AGC CCT TTC TCT CCC
 1248
 Ala Thr Gly Glu Ala Val Leu Tyr Glu Ile Ser Ser Pro Phe Ser Pro
 405 410 415

ATA ATG GAT CCC CTA CCA ATA CGC ACC AAA AGC AAC ACT AGC AGG AAA
 1296
 Ile Met Asp Pro Leu Pro Ile Arg Thr Lys Ser Asn Thr Ser Arg Lys
 420 425 430

GAC TGG GCT CCC CAG TCA ACC CCA AGT AAG GAT TCT TTC CAC CCC AGT
 1344
 Asp Trp Ala Pro Gln Ser Thr Pro Ser Lys Asp Ser Phe His Pro Ser
 435 440 445

TCT CTT ATG AGT GCC CTC ATC CAG CAG GAT GAG TCC ATC TAT CTG TGT
 1392
 Ser Leu Met Ser Ala Leu Ile Gln Gln Asp Glu Ser Ile Tyr Leu Cys
 450 455 460

CCT CCT TCA AGC CCT GCG CTG TTA GAC AGC CAT TTT CTC ATG GGC TCC
 1440

0055402.05139

Pro Pro Ser Ser Pro Ala Leu Leu Asp Ser His Phe Leu Met Gly Ser
465 470 475 480

GTG AGC AAG TGC GGG AGT TGG CAA GAC AGC TTT GCG GCC GCA GGA AGT
1488

Val Ser Lys Cys Gly Ser Trp Gln Asp Ser Phe Ala Ala Ala Gly Ser
485 490 495

GAG GCT GCG CTG AAA CAT GAG CAA ATT GGC CAT GCT CAG GAC GTG AAC
1536

Glu Ala Ala Leu Lys His Glu Gln Ile Gly His Ala Gln Asp Val Asn
500 505 510

CTT GCA CTC TCT GGC GGC CCC TCA GAG CTC TTT CCG GAT AAT AAA AAT
1584

Leu Ala Leu Ser Gly Gly Pro Ser Glu Leu Phe Pro Asp Asn Lys Asn
515 520 525

AAT GAC TTG TAC AGC ATC ATG AGG AAC CTT GGG ATT GAT TTT GAA GAT
1632

Asn Asp Leu Tyr Ser Ile Met Arg Asn Leu Gly Ile Asp Phe Glu Asp
530 535 540

ATC AGA AGC ATG CAG AAC GAG GAG TTC TTC AGA ACT GAC TCC ACC GCT
1680

Ile Arg Ser Met Gln Asn Glu Glu Phe Phe Arg Thr Asp Ser Thr Ala
545 550 555 560

GCT GGT GAG GTT GAC TTC AAA GAC ATC GAC ATA ACG GAC GAA ATC CTG
1728

Ala Gly Glu Val Asp Phe Lys Asp Ile Asp Ile Thr Asp Glu Ile Leu
565 570 575

ACC TAC GTG CAG GAT TCC CTG AAC AAT TCA ACT TTG CTG AAC TCG GCT
1776

Thr Tyr Val Gln Asp Ser Leu Asn Asn Ser Thr Leu Leu Asn Ser Ala
580 585 590

TGC CAG CAG CAG CCT GTG ACT CAG CAC CTA AGC TGT ATG CTG CAG GAG
1824

Cys Gln Gln Gln Pro Val Thr Gln His Leu Ser Cys Met Leu Gln Glu
595 600 605

CGC CTG CAA CTA GAG CAA CAG CAA CAG CTT CAG CAG CCC CCG CCG CAG
1872

Arg Leu Gln Leu Glu Gln Gln Gln Gln Leu Gln Gln Pro Pro Pro Gln
610 615 620

GCT CTG GAG CCC CAG CAG CAG CTG TGT CAG ATG GTG TGC CCC CAG CAA
1920

Ala Leu Glu Pro Gln Gln Gln Leu Cys Gln Met Val Cys Pro Gln Gln
625 630 635 640

0985402-051397

GAT CTG GGT CCG AAG CAC ACG CAA ATC AAC GGC ACG TTT GCA AGT TGG
1968

Asp Leu Gly Pro Lys His Thr Gln Ile Asn Gly Thr Phe Ala Ser Trp
645 650 655

AAC CCC ACC CCT CCC GTG TCT TTC AAC TGT CCC CAG CAG GAA CTA AAG
2016

Asn Pro Thr Pro Pro Val Ser Phe Asn Cys Pro Gln Gln Glu Leu Lys
660 665 670

CAC TAT CAG CTC TTT TCC AGC TTA CAG GGG ACT GCT CAG GAA TTT CCC
2064

His Tyr Gln Leu Phe Ser Ser Leu Gln Gly Thr Ala Gln Glu Phe Pro
675 680 685

TAC AAA CCA GAG GTG GAC AGT GTG CCT TAC ACA CAG AAC TTT GCT CCC
2112

Tyr Lys Pro Glu Val Asp Ser Val Pro Tyr Thr Gln Asn Phe Ala Pro
690 695 700

TGT AAT CAG CCT CTG CTT CCA GAA CAT TCC AAG AGT GTG CAG TTG GAC
2160

Cys Asn Gln Pro Leu Leu Pro Glu His Ser Lys Ser Val Gln Leu Asp
705 710 715 720

TTC CCT GGA AGG GAT TTT GAA CCG TCC CTG CAT CCC ACT ACT TCT AAT
2208

Phe Pro Gly Arg Asp Phe Glu Pro Ser Leu His Pro Thr Thr Ser Asn
725 730 735

TTA GAT TTT GTC AGT TGT TTA CAA GTT CCT GAA AAC CAA AGT CAT GGG
2256

Leu Asp Phe Val Ser Cys Leu Gln Val Pro Glu Asn Gln Ser His Gly
740 745 750

ATA AAC TCA CAG TCC GCC ATG GTC AGT CCT CAG GCA TAC TAT GCT GGG
2304

Ile Asn Ser Gln Ser Ala Met Val Ser Pro Gln Ala Tyr Tyr Ala Gly
755 760 765

GCC ATG TCC ATG TAT CAG TGC CAG CCA GGG CCA CAG CGC ACC CCT GTG
2352

Ala Met Ser Met Tyr Gln Cys Gln Pro Gly Pro Gln Arg Thr Pro Val
770 775 780

GAC CAG ACG CAG TAC AGC TCT GAA ATT CCA GGT TCT CAG GCA TTC CTA
2400

Asp Gln Thr Gln Tyr Ser Ser Glu Ile Pro Gly Ser Gln Ala Phe Leu
785 790 795 800

AGC AAG GTG CAG AGT TGAGGTGTTT TCAATGAAAC CTATTCGTCC GACTTGAGCA
2455

0005402-051397

Ser Lys Val Gln Ser
805

GCATTGGCCA CGCTGCTCAG ACCACTGGCC ATCTCCATCA CTGCGGAAGC CCGGCCTCTT
2515

CCCGATATCA CACCCGGTGG ATTCCTGTAG CTCCCATGCC AGGATGAAAT TCATTTCAGGA
2575

ACAGGATACC AGAACTGTGA GGGTTGGACA TCAGTACACT TTCTCCAAAA CAGATTTCTGA
2635

TTCTTGTGTT TAGAGAAGGA GTTTAAAACC CGTACCTGAG ATGCTCCCTA TACGATGGGA
2695

GAGCTCGGAC GGAGCACATG GGAGGAGTTC AGGCACCTCA GAGTGCACAG TGTTTACTGT
2755

GAAAAATTCT CGGGTTCCT GCTCAGTAAC TTCAGCAGGA AAAACAGGGA GGTATTTGGA
2815

GCTTTGAACT TCTGGATTCT TGTTAGTATA CCAAATACGG AGTTACAGGA CTAACCGATT
2875

TCCTATATTT TTTAACCTCT GTTTTTGTCC CAGAAGTTAA AGTAAATGGT TTGGTGCTTT
2935

TCTCAAAAGA AAATCTCAAT GCTTTCTTTC TGCACTGTTA ATATAAGTGC CTCACTTTTT
2995

GTGTTGTTG TTGTTGTTTT CTGATTTTTT TCTTTTTTTC TATCTACCTG TAACACAATA
3055

GGGTATGTAT TTTATATGAA ATATTTTTTA TCTTTTTTGA ATTAATATTC TTTCTGCACA
3115

AAGAAAGTTT CCCGAATCCC AACCTTTCTA TGACCCCGCT GTGTGTGTGC ACTACTCATC
3175

CTTTCCTTCA GATAAAGAGT AATTGATAAC TC
3207

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 805 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

0855402.051397

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ser Ser Gly Ala Asn Ile Thr Tyr Ala Ser Arg Lys Arg Arg Lys
 1 5 10 15
 Pro Val Gln Lys Thr Val Lys Pro Ile Pro Ala Glu Gly Ile Lys Ser
 20 25 30
 Asn Pro Ser Lys Arg His Arg Asp Arg Leu Asn Thr Glu Leu Asp Arg
 35 40 45
 Leu Ala Ser Leu Leu Pro Phe Pro Gln Asp Val Ile Asn Lys Leu Asp
 50 55 60
 Lys Leu Ser Val Leu Arg Leu Ser Val Thr Tyr Leu Arg Ala Lys Ser
 65 70 75 80
 Phe Phe Asp Val Ala Leu Lys Ser Thr Pro Ala Asp Arg Asn Gly Gly
 85 90 95
 Gln Asp Gln Cys Arg Ala Gln Ile Arg Asp Trp Gln Asp Leu Gln Glu
 100 105 110
 Gly Glu Phe Leu Leu Gln Ala Leu Asn Gly Phe Val Leu Val Val Thr
 115 120 125
 Ala Asp Ala Leu Val Phe Tyr Ala Ser Ser Thr Ile Gln Asp Tyr Leu
 130 135 140
 Gly Phe Gln Gln Ser Asp Val Ile His Gln Ser Val Tyr Glu Leu Ile
 145 150 155 160
 His Thr Glu Asp Arg Ala Glu Phe Gln Arg Gln Leu His Trp Ala Leu
 165 170 175
 Asn Pro Asp Ser Ala Gln Gly Val Asp Glu Ala His Gly Pro Pro Gln
 180 185 190
 Ala Ala Val Tyr Tyr Thr Pro Asp Gln Leu Pro Pro Glu Asn Ala Ser
 195 200 205
 Phe Met Glu Arg Cys Phe Arg Cys Arg Leu Arg Cys Leu Leu Asp Asn
 210 215 220
 Ser Ser Gly Phe Leu Ala Met Asn Phe Gln Gly Arg Leu Lys Tyr Leu
 225 230 235 240
 His Gly Gln Asn Lys Lys Gly Lys Asp Gly Ala Leu Leu Pro Pro Gln
 245 250 255
 Leu Ala Leu Phe Ala Ile Ala Thr Pro Leu Gln Pro Pro Ser Ile Leu
 260 265 270

201501051397

Glu	Ile	Arg	Thr	Lys	Asn	Phe	Ile	Phe	Arg	Thr	Lys	His	Lys	Leu	Asp
		275					280					285			
Phe	Thr	Pro	Ile	Gly	Cys	Asp	Ala	Lys	Gly	Gln	Leu	Ile	Leu	Gly	Tyr
	290					295					300				
Thr	Glu	Val	Glu	Leu	Cys	Thr	Arg	Gly	Ser	Gly	Tyr	Gln	Phe	Ile	His
305					310					315					320
Ala	Ala	Asp	Ile	Leu	His	Cys	Ala	Glu	Ser	His	Ile	Arg	Met	Ile	Lys
				325					330					335	
Thr	Gly	Glu	Ser	Gly	Met	Thr	Val	Phe	Arg	Leu	Leu	Ala	Lys	His	Ser
			340					345					350		
Arg	Trp	Arg	Trp	Val	Gln	Ser	Asn	Ala	Arg	Leu	Ile	Tyr	Arg	Asn	Gly
		355					360					365			
Arg	Pro	Asp	Tyr	Ile	Ile	Ala	Thr	Gln	Arg	Pro	Leu	Thr	Asp	Glu	Glu
	370					375					380				
Gly	Arg	Glu	His	Leu	Gln	Lys	Arg	Ser	Thr	Ser	Leu	Pro	Phe	Met	Phe
385					390					395					400
Ala	Thr	Gly	Glu	Ala	Val	Leu	Tyr	Glu	Ile	Ser	Ser	Pro	Phe	Ser	Pro
				405					410					415	
Ile	Met	Asp	Pro	Leu	Pro	Ile	Arg	Thr	Lys	Ser	Asn	Thr	Ser	Arg	Lys
			420					425					430		
Asp	Trp	Ala	Pro	Gln	Ser	Thr	Pro	Ser	Lys	Asp	Ser	Phe	His	Pro	Ser
		435					440					445			
Ser	Leu	Met	Ser	Ala	Leu	Ile	Gln	Gln	Asp	Glu	Ser	Ile	Tyr	Leu	Cys
	450					455					460				
Pro	Pro	Ser	Ser	Pro	Ala	Leu	Leu	Asp	Ser	His	Phe	Leu	Met	Gly	Ser
465					470					475					480
Val	Ser	Lys	Cys	Gly	Ser	Trp	Gln	Asp	Ser	Phe	Ala	Ala	Ala	Gly	Ser
				485					490					495	
Glu	Ala	Ala	Leu	Lys	His	Glu	Gln	Ile	Gly	His	Ala	Gln	Asp	Val	Asn
			500					505					510		
Leu	Ala	Leu	Ser	Gly	Gly	Pro	Ser	Glu	Leu	Phe	Pro	Asp	Asn	Lys	Asn
		515					520					525			
Asn	Asp	Leu	Tyr	Ser	Ile	Met	Arg	Asn	Leu	Gly	Ile	Asp	Phe	Glu	Asp
	530					535					540				
Ile	Arg	Ser	Met	Gln	Asn	Glu	Glu	Phe	Phe	Arg	Thr	Asp	Ser	Thr	Ala

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5261 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 383..2927

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AATTCGCAC GGCCAGACC CAGGATTCTT TATAGACGGC CCAGGCTCCT CCTCCGCCCG
60

GGCCGCCTCA CCTGCGGGCA TTGCGCGCCG CCTCCGCCCG TGTAGACGGC ACCTGCGCCG
120

CCTTGCTCGC GGGTCTCCGC CCTCGCCAC CCTCACTGCG CCAGGCCAG GCAGCTCACC
180

TGTACTGGCG CGGGCTGCGG AAGCTGCGTG ACGCGAGGCG TTGAGGCGCG GCGCCACGC
240

CACTGTCCCG AGAGGACGCA GGTGGAGCGG GCGCGACTTC GCGAACCCGG CGCCGGCCGC
300

CGCAGTGGTC CCAGCCTACA CCGGGTTCCG GGGACCCGGC CGCCAGTGCC CGGGGAGTAG
360

CCGCCGCCGT CGGCTGGGCA CC ATG AAC AGC AGC AGC GCC AAC ATC ACC TAC
412

Met Asn Ser Ser Ser Ala Asn Ile Thr Tyr

1

5

10

GCC AGT CGC AAG CGG CGG AAG CCG GTG CAG AAA ACA GTA AAG CCA ATC
460

Ala Ser Arg Lys Arg Arg Lys Pro Val Gln Lys Thr Val Lys Pro Ile
15 20 25

CCA GCT GAA GGA ATC AAG TCA AAT CCT TCC AAG CGG CAT AGA GAC CGA
508

Pro Ala Glu Gly Ile Lys Ser Asn Pro Ser Lys Arg His Arg Asp Arg
30 35 40

CTT AAT ACA GAG TTG GAC CGT TTG GCT AGC CTG CTG CCT TTC CCA CAA

20150105199

Leu	Asn	Thr	Glu	Leu	Asp	Arg	Leu	Ala	Ser	Leu	Leu	Pro	Phe	Pro	Gln
		45					50					55			
GAT	GTT	ATT	AAT	AAG	TTG	GAC	AAA	CTT	TCA	GTT	CTT	AGG	CTC	AGC	GTC
604															
Asp	Val	Ile	Asn	Lys	Leu	Asp	Lys	Leu	Ser	Val	Leu	Arg	Leu	Ser	Val
60						65					70				
AGT	TAC	CTG	AGA	GCC	AAG	AGC	TTC	TTT	GAT	GTT	GCA	TTA	AAA	TCC	TCC
652															
Ser	Tyr	Leu	Arg	Ala	Lys	Ser	Phe	Phe	Asp	Val	Ala	Leu	Lys	Ser	Ser
75					80					85					90
CCT	ACT	GAA	AGA	AAC	GGA	GGC	CAG	GAT	AAC	TGT	AGA	GCA	GCA	AAT	TTC
700															
Pro	Thr	Glu	Arg	Asn	Gly	Gly	Gln	Asp	Asn	Cys	Arg	Ala	Ala	Asn	Phe
				95					100					105	
AGA	GAA	GGC	CTG	AAC	TTA	CAA	GAA	GGA	GAA	TTC	TTA	TTA	CAG	GCT	CTG
748															
Arg	Glu	Gly	Leu	Asn	Leu	Gln	Glu	Gly	Glu	Phe	Leu	Leu	Gln	Ala	Leu
			110					115					120		
AAT	GGC	TTT	GTA	TTA	GTT	GTC	ACT	ACA	GAT	GCT	TTG	GTC	TTT	TAT	GCT
796															
Asn	Gly	Phe	Val	Leu	Val	Val	Thr	Thr	Asp	Ala	Leu	Val	Phe	Tyr	Ala
		125					130					135			
TCT	TCT	ACT	ATA	CAA	GAT	TAT	CTA	GGG	TTT	CAG	CAG	TCT	GAT	GTC	ATA
844															
Ser	Ser	Thr	Ile	Gln	Asp	Tyr	Leu	Gly	Phe	Gln	Gln	Ser	Asp	Val	Ile
140						145					150				
CAT	CAG	AGT	GTA	TAT	GAA	CTT	ATC	CAT	ACC	GAA	GAC	CGA	GCT	GAA	TTT
892															
His	Gln	Ser	Val	Tyr	Glu	Leu	Ile	His	Thr	Glu	Asp	Arg	Ala	Glu	Phe
155					160					165					170
CAG	CGT	CAG	CTA	CAC	TGG	GCA	TTA	AAT	CCT	TCT	CAG	TGT	ACA	GAG	TCT
940															
Gln	Arg	Gln	Leu	His	Trp	Ala	Leu	Asn	Pro	Ser	Gln	Cys	Thr	Glu	Ser
				175					180					185	
GGA	CAA	GGA	ATT	GAA	GAA	GCC	ACT	GGT	CTC	CCC	CAG	ACA	GTA	GTC	TGT
988															
Gly	Gln	Gly	Ile	Glu	Glu	Ala	Thr	Gly	Leu	Pro	Gln	Thr	Val	Val	Cys
			190					195					200		
TAT	AAC	CCA	GAC	CAG	ATT	CCT	CCA	GAA	AAC	TCT	CCT	TTA	ATG	GAG	AGG
1036															
Tyr	Asn	Pro	Asp	Gln	Ile	Pro	Pro	Glu	Asn	Ser	Pro	Leu	Met	Glu	Arg

205				210				215							
TGC	TTC	ATA	TGT	CGT	CTA	AGG	TGT	CTG	CTG	GAT	AAT	TCA	TCT	GGT	TTT
1084															
Cys	Phe	Ile	Cys	Arg	Leu	Arg	Cys	Leu	Leu	Asp	Asn	Ser	Ser	Gly	Phe
220						225					230				
CTG	GCA	ATG	AAT	TTC	CAA	GGG	AAG	TTA	AAG	TAT	CTT	CAT	GGA	CAG	AAA
1132															
Leu	Ala	Met	Asn	Phe	Gln	Gly	Lys	Leu	Lys	Tyr	Leu	His	Gly	Gln	Lys
235					240					245					250
AAG	AAA	GGG	AAA	GAT	GGA	TCA	ATA	CTT	CCA	CCT	CAG	TTG	GCT	TTG	TTT
1180															
Lys	Lys	Gly	Lys	Asp	Gly	Ser	Ile	Leu	Pro	Pro	Gln	Leu	Ala	Leu	Phe
				255					260					265	
GCG	ATA	GCT	ACT	CCA	CTT	CAG	CCA	CCA	TCC	ATA	CTT	GAA	ATC	CGG	ACC
1228															
Ala	Ile	Ala	Thr	Pro	Leu	Gln	Pro	Pro	Ser	Ile	Leu	Glu	Ile	Arg	Thr
			270						275					280	
AAA	AAT	TTT	ATC	TTT	AGA	ACC	AAA	CAC	AAA	CTA	GAC	TTC	ACA	CCT	ATT
1276															
Lys	Asn	Phe	Ile	Phe	Arg	Thr	Lys	His	Lys	Leu	Asp	Phe	Thr	Pro	Ile
		285					290					295			
GGT	TGT	GAT	GCC	AAA	GGA	AGA	ATT	GTT	TTA	GGA	TAT	ACT	GAA	GCA	GAG
1324															
Gly	Cys	Asp	Ala	Lys	Gly	Arg	Ile	Val	Leu	Gly	Tyr	Thr	Glu	Ala	Glu
300						305					310				
CTG	TGC	ACG	AGA	GGC	TCA	GGT	TAT	CAG	TTT	ATT	CAT	GCA	GCT	GAT	ATG
1372															
Leu	Cys	Thr	Arg	Gly	Ser	Gly	Tyr	Gln	Phe	Ile	His	Ala	Ala	Asp	Met
315					320					325					330
CTT	TAT	TGT	GCC	GAG	TCC	CAT	ATC	CGA	ATG	ATT	AAG	ACT	GGA	GAA	AGT
1420															
Leu	Tyr	Cys	Ala	Glu	Ser	His	Ile	Arg	Met	Ile	Lys	Thr	Gly	Glu	Ser
				335					340					345	
GGC	ATG	ATA	GTT	TTC	CGG	CTT	CTT	ACA	AAA	AAC	AAC	CGA	TGG	ACT	TGG
1468															
Gly	Met	Ile	Val	Phe	Arg	Leu	Leu	Thr	Lys	Asn	Asn	Arg	Trp	Thr	Trp
			350						355				360		
GTC	CAG	TCT	AAT	GCA	CGC	CTG	CTT	TAT	AAA	AAT	GGA	AGA	CCA	GAT	TAT
1516															
Val	Gln	Ser	Asn	Ala	Arg	Leu	Leu	Tyr	Lys	Asn	Gly	Arg	Pro	Asp	Tyr
			365				370							375	

08855402-051397

ATC ATT GTA ACT CAG AGA CCA CTA ACA GAT GAG GAA GGA ACA GAG CAT
 1564
 Ile Ile Val Thr Gln Arg Pro Leu Thr Asp Glu Glu Gly Thr Glu His
 380 385 390

TTA CGA AAA CGA AAT ACG AAG TTG CCT TTT ATG TTT ACC ACT GGA GAA
 1612
 Leu Arg Lys Arg Asn Thr Lys Leu Pro Phe Met Phe Thr Thr Gly Glu
 395 400 405 410

GCT GTG TTG TAT GAG GCA ACC AAC CCT TTT CCT GCC ATA ATG GAT CCC
 1660
 Ala Val Leu Tyr Glu Ala Thr Asn Pro Phe Pro Ala Ile Met Asp Pro
 415 420 425

TTA CCA CTA AGG ACT AAA AAT GGC ACT AGT GGA AAA GAC TCT GCT ACC
 1708
 Leu Pro Leu Arg Thr Lys Asn Gly Thr Ser Gly Lys Asp Ser Ala Thr
 430 435 440

ACA TCC ACT CTA AGC AAG GAC TCT CTC AAT CCT AGT TCC CTC CTG GCT
 1756
 Thr Ser Thr Leu Ser Lys Asp Ser Leu Asn Pro Ser Ser Leu Leu Ala
 445 450 455

GCC ATG ATG CAA CAA GAT GAG TCT ATT TAT CTC TAT CCT GCT TCA AGT
 1804
 Ala Met Met Gln Gln Asp Glu Ser Ile Tyr Leu Tyr Pro Ala Ser Ser
 460 465 470

ACT TCA AGT ACT GCA CCT TTT GAA AAC AAC TTT TTC AAC GAA TCT ATG
 1852
 Thr Ser Ser Thr Ala Pro Phe Glu Asn Asn Phe Phe Asn Glu Ser Met
 475 480 485 490

AAT GAA TGC AGA AAT TGG CAA GAT AAT ACT GCA CCG ATG GGA AAT GAT
 1900
 Asn Glu Cys Arg Asn Trp Gln Asp Asn Thr Ala Pro Met Gly Asn Asp
 495 500 505

ACT ATC CTG AAA CAT GAG CAA ATT GAC CAG CCT CAG GAT GTG AAC TCA
 1948
 Thr Ile Leu Lys His Glu Gln Ile Asp Gln Pro Gln Asp Val Asn Ser
 510 515 520

TTT GCT GGA GGT CAC CCA GGG CTC TTT CAA GAT AGT AAA AAC AGT GAC
 1996
 Phe Ala Gly Gly His Pro Gly Leu Phe Gln Asp Ser Lys Asn Ser Asp
 525 530 535

TTG TAC AGC ATA ATG AAA AAC CTA GGC ATT GAT TTT GAA GAC ATC AGA
 2044

2025-10-20 15:55:50

Asn Phe Ile Ser Cys Asn Gln Pro Val Leu Pro Gln His Ser Lys Cys
700 705 710

ACA GAG CTG GAC TAC CCT ATG GGG AGT TTT GAA CCA TCC CCA TAC CCC
 2572
 Thr Glu Leu Asp Tyr Pro Met Gly Ser Phe Glu Pro Ser Pro Tyr Pro
 715 720 725 730

ACT ACT TCT AGT TTA GAA GAT TTT GTC ACT TGT TTA CAA CTT CCT GAA
 2620
 Thr Thr Ser Ser Leu Glu Asp Phe Val Thr Cys Leu Gln Leu Pro Glu
 735 740 745

AAC CAA AAG CAT GGA TTA AAT CCA CAG TCA GCC ATA ATA ACT CCT CAG
 2668
 Asn Gln Lys His Gly Leu Asn Pro Gln Ser Ala Ile Ile Thr Pro Gln
 750 755 760

ACA TGT TAT GCT GGG GCC GTG TCG ATG TAT CAG TGC CAG CCA GAA CCT
 2716
 Thr Cys Tyr Ala Gly Ala Val Ser Met Tyr Gln Cys Gln Pro Glu Pro
 765 770 775

CAG CAC ACC CAC GTG GGT CAG ATG CAG TAC AAT CCA GTA CTG CCA GGC
 2764
 Gln His Thr His Val Gly Gln Met Gln Tyr Asn Pro Val Leu Pro Gly
 780 785 790

CAA CAG GCA TTT TTA AAC AAG TTT CAG AAT GGA GTT TTA AAT GAA ACA
 2812
 Gln Gln Ala Phe Leu Asn Lys Phe Gln Asn Gly Val Leu Asn Glu Thr
 795 800 805 810

TAT CCA GCT GAA TTA AAT AAC ATA AAT AAC ACT CAG ACT ACC ACA CAT
 2860
 Tyr Pro Ala Glu Leu Asn Asn Ile Asn Asn Thr Gln Thr Thr Thr His
 815 820 825

CTT CAG CCA CTT CAT CAT CCG TCA GAA GCC AGA CCT TTT CCT GAT TTG
 2908
 Leu Gln Pro Leu His His Pro Ser Glu Ala Arg Pro Phe Pro Asp Leu
 830 835 840

ACA TCC AGT GGA TTC CTG T AATTCCAAGC CCAATTTTGA CCCTGGTTTT
 2957
 Thr Ser Ser Gly Phe Leu
 845

TGGATTAAAT TAGTTTGTGA AGGATTATGG AAAAATAAAA CTGTCACTGT TGGACGTCAG
 3017

CAAGTTCACA TGGAGGCATT GATGCATGCT ATTCACAATT ATTCCAAACC AAATTTTAAT
 3077

TTTTGCTTTT AGAAAAGGGA GTTTAAAAAT GGTATCAAAA TTACATATAC TACAGTCAAG

2045580-05139

3137

ATAGAAAGGG TGCTGCCACG GAGTGGTGAG GTACCGTCTA CATTTACAT TATTCTGGGC
3197

ACCACAAAAT ATACAAAAC TATCAGGGA AACTAAGATT CTTTTAAATT AGAAAATATT
3257

CTCTATTTGA ATTATTTCTG TCACAGTAAA AATAAAATAC TTTGAGTTTT GAGCTACTGG
3317

ATTCTTATTA GTTCCCCAAA TACAAAGTTA GAGAACTAAA CTAGTTTTTC CTATCATGTT
3377

AACCTCTGCT TTTATCTCAG ATGTTAAAT AAATGGTTTG GTGCTTTTTA TAAAAAGATA
3437

ATCTCAGTGC TTTCCTCCTT CACTGTTTCA TCTAAGTGCC TCACATTTTT TTCTACCTAT
3497

AACACTCTAG GATGTATATT TTATATAAAG TATTCTTTTT CTTTTTAAA TTAATATCTT
3557

TCTGCACACA AATATTATTT GTGTTTCCTA AATCCAACCA ATTTTCATTA ATTCAGGCAT
3617

ATTTTAACTC CACTGCTTAC CTACTTTCTT CAGGTAAAAG GGCAAATAAT GATCGAAAAA
3677

ATAATTATTT ATTACATAAT TTAGTTGTTT CTAGACTATA AATGTTGCTA TGTGCCTTAT
3737

GTTGAAAAAA TTTAAAAGTA AAATGTCTTT CCAAATTATT TCTTAATTAT TATAAAAATA
3797

TTAAGACAAT AGCACTTAAA TTCCTCAACA GTGTTTTTCAG AAGAAATAAA TATACCACTC
3857

TTTACCTTTA TTGATATCTC CATGATGATA GTTGAATGTT GCAATGTGAA AAATCTGCTG
3917

TTAACTGCAA CCTTGTTTAT TAAATTGCAA GAAGCTTTAT TTCTAGCTTT TTAATTAAGC
3977

AAAGCACCCA TTTCAATGTG TATAAATTGT CTTTAAAAAC TGTTTTAGAC CTATAATCCT
4037

TGATAATATA TTGTGTTGAC TTTATAAATT TCGCTTCTTA GAACAGTGGA AACTATGTGT
4097

TTTTCTCATA TTTGAGGAGT GTTAAGATTG CAGATAGCAA GGTTTGGTGC AAAGTATTGT
4157

0855402 051397
26ET50 20455280

AATGAGTGAA 4217	TTGAATGGTG	CATTGTATAG	ATATAATGAA	CAAAATTATT	TGTAAGATAT
TTGCAGTTTT 4277	TCATTTTAAA	AAGTCCATAC	CTTATATATG	CACTTAATTT	GTTGGGGCTT
TACATACTTT 4337	ATCAATGTGT	CTTTCTAAGA	AATCAAGTAA	TGAATCCAAC	TGCTTAAAGT
TGGTATTAAT 4397	AAAAAGACAA	CCACATAGTT	CGTTTACCTT	CAAACCTTAG	GTTTTTTTAA
TGATATACTG 4457	ATCTTCATTA	CCAATAGGCA	AATTAATCAC	CCTACCAACT	TTACTGTCCT
AACATGGACT 4517	TTCAAAAAGA	AAAAATGACA	CCATCTTTTA	TTCTTTTTTTT	TTTTTTTTTTT
TTGAGAGAGA 4577	GTCTTACTCT	GCCGCCCAAA	CTGGAGTGCA	GTGGCACAAT	CTTGGCTCAC
TGCAACCTCT 4637	ACCTCCTGGG	TTCAAGTGAT	TCTCTTGCCCT	CAGCCTCCCG	AGTTGCTGGG
ATTGCGGGCA 4697	TGGTGGCGTG	AGCCTGTAGT	CCTAGCTACT	CGGGAGGCTG	AGGCAGGAGA
ATAGCCTGAA 4757	CCTGGGAATC	GGAGGTTGCA	GGGCCAAGAT	CGCCCCACTG	CACTCCAGCC
TGGCAATAGA 4817	CCGAGCTCCG	TCTCCAAAAA	AAAAAATACA	ATTTTTATTT	CTTTTACTTT
TTTTAGTAAG 4877	TTAATGTATA	TAAAAATGGC	TTCGGACAAA	ATATCTCTGA	GTTCTGTGTA
TTTTCAGTCA 4937	AAACTTTAAA	CCTGTAGAAT	CAATTTAAGT	GTTGAAAAAA	ATTTGTCTGA
AACATTTTCAT 4997	AATTTGTTTC	CAGCATGAGG	TATCTAAGGA	TTTAGACCAG	AGGTCTAGAT
TAATACTCTA 5057	TTTTTACATT	TAAACCTTTT	ATTATAAGTC	TTACATAAAC	CATTTTTGTT
ACTCTCTTCC 5117	ACATGTTACT	GGATAAATTG	TTTAGTGGA	AATAGGCTTT	TTAATCATGA
ATATGATGAC 5177	AATCAGTTAT	ACAGTTATAA	AATTAAAAGT	TTGAAAAGCA	ATATTGTATA
TTTTTATCTA	TATAAAATAA	CTAAAATGTA	TCTAAGAATA	ATAAAAATCAC	GTAAACCAA

5237

AAAAAAAAAA AAAAAAAAAA AAAA

5261

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 848 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Asn Ser Ser Ser Ala Asn Ile Thr Tyr Ala Ser Arg Lys Arg Arg
 1 5 10 15

Lys Pro Val Gln Lys Thr Val Lys Pro Ile Pro Ala Glu Gly Ile Lys
 20 25 30

Ser Asn Pro Ser Lys Arg His Arg Asp Arg Leu Asn Thr Glu Leu Asp
 35 40 45

Arg Leu Ala Ser Leu Leu Pro Phe Pro Gln Asp Val Ile Asn Lys Leu
 50 55 60

Asp Lys Leu Ser Val Leu Arg Leu Ser Val Ser Tyr Leu Arg Ala Lys
 65 70 75 80

Ser Phe Phe Asp Val Ala Leu Lys Ser Ser Pro Thr Glu Arg Asn Gly
 85 90 95

Gly Gln Asp Asn Cys Arg Ala Ala Asn Phe Arg Glu Gly Leu Asn Leu
 100 105 110

Gln Glu Gly Glu Phe Leu Leu Gln Ala Leu Asn Gly Phe Val Leu Val
 115 120 125

Val Thr Thr Asp Ala Leu Val Phe Tyr Ala Ser Ser Thr Ile Gln Asp
 130 135 140

Tyr Leu Gly Phe Gln Gln Ser Asp Val Ile His Gln Ser Val Tyr Glu
 145 150 155 160

Leu Ile His Thr Glu Asp Arg Ala Glu Phe Gln Arg Gln Leu His Trp
 165 170 175

Ala Leu Asn Pro Ser Gln Cys Thr Glu Ser Gly Gln Gly Ile Glu Glu
 180 185 190

20150520 15:40:02

Ala Thr Gly Leu Pro Gln Thr Val Val Cys Tyr Asn Pro Asp Gln Ile
195 200 205

Pro Pro Glu Asn Ser Pro Leu Met Glu Arg Cys Phe Ile Cys Arg Leu
210 215 220

Arg Cys Leu Leu Asp Asn Ser Ser Gly Phe Leu Ala Met Asn Phe Gln
225 230 235 240

Gly Lys Leu Lys Tyr Leu His Gly Gln Lys Lys Lys Gly Lys Asp Gly
245 250 255

Ser Ile Leu Pro Pro Gln Leu Ala Leu Phe Ala Ile Ala Thr Pro Leu
260 265 270

Gln Pro Pro Ser Ile Leu Glu Ile Arg Thr Lys Asn Phe Ile Phe Arg
275 280 285

Thr Lys His Lys Leu Asp Phe Thr Pro Ile Gly Cys Asp Ala Lys Gly
290 295 300

Arg Ile Val Leu Gly Tyr Thr Glu Ala Glu Leu Cys Thr Arg Gly Ser
305 310 315 320

Gly Tyr Gln Phe Ile His Ala Ala Asp Met Leu Tyr Cys Ala Glu Ser
325 330 335

His Ile Arg Met Ile Lys Thr Gly Glu Ser Gly Met Ile Val Phe Arg
340 345 350

Leu Leu Thr Lys Asn Asn Arg Trp Thr Trp Val Gln Ser Asn Ala Arg
355 360 365

Leu Leu Tyr Lys Asn Gly Arg Pro Asp Tyr Ile Ile Val Thr Gln Arg
370 375 380

Pro Leu Thr Asp Glu Glu Gly Thr Glu His Leu Arg Lys Arg Asn Thr
385 390 395 400

Lys Leu Pro Phe Met Phe Thr Thr Gly Glu Ala Val Leu Tyr Glu Ala
405 410 415

Thr Asn Pro Phe Pro Ala Ile Met Asp Pro Leu Pro Leu Arg Thr Lys
420 425 430

Asn Gly Thr Ser Gly Lys Asp Ser Ala Thr Thr Ser Thr Leu Ser Lys
435 440 445

Asp Ser Leu Asn Pro Ser Ser Leu Leu Ala Ala Met Met Gln Gln Asp
450 455 460

Glu Ser Ile Tyr Leu Tyr Pro Ala Ser Ser Thr Ser Ser Thr Ala Pro

25540-2045580

Asn	Pro	Gln	Ser	Ala	Ile	Ile	Thr	Pro	Gln	Thr	Cys	Tyr	Ala	Gly	Ala
		755					760					765			
Val	Ser	Met	Tyr	Gln	Cys	Gln	Pro	Glu	Pro	Gln	His	Thr	His	Val	Gly
	770					775					780				
Gln	Met	Gln	Tyr	Asn	Pro	Val	Leu	Pro	Gly	Gln	Gln	Ala	Phe	Leu	Asn
785					790					795					800
Lys	Phe	Gln	Asn	Gly	Val	Leu	Asn	Glu	Thr	Tyr	Pro	Ala	Glu	Leu	Asn
				805					810					815	
Asn	Ile	Asn	Asn	Thr	Gln	Thr	Thr	Thr	His	Leu	Gln	Pro	Leu	His	His
			820					825					830		
Pro	Ser	Glu	Ala	Arg	Pro	Phe	Pro	Asp	Leu	Thr	Ser	Ser	Gly	Phe	Leu
		835					840					845			

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 6
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 13
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base

264550-2045500

- (B) LOCATION: 19
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 22
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 29
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 32
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 43
- (D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 46
- (D) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TTNATNCCTC TCNGCNGGNA TNGGTCTTNA CNGTTCTTTC TGNACNGGTC TT
52

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 7
- (D) OTHER INFORMATION: /mod_base= OTHER
/note= "Can be either adenine, thymine,
guanosine,
or cytosine."

2025-10-27 15:13:07

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 10
- (D) OTHER INFORMATION: /mod_base= OTHER
/note= "Can be either adenine, thymine,
guanosine,
or cytosine."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AAAGCCNGTN CAAGAAAGAC
20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGATTGACT TAATTCCTTC AGGGG
25

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCATCGATCT CGAGAGATTG CAGATAGCAA GGTTTGGTGC
40

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs

25055402.051397

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCATCGATCT CGAGTGTAAT GAGTGAATTG AATGGTGC
38

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAAGATCTTC CAGTGGTCCC AGCCTACACC
30

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAAGATCTTC ATGTGAACTT GCTGACGTCC
30

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 63 base pairs
- (B) TYPE: nucleic acid

26250-20455330

- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCTCTAGATG ATCACCATGG TGCAGAAGAC CGTGAAGCCC ATCCCCGCTG AAGGAATTAA
60

GTC

63

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 63 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCACTAGTTG ATCACCATGG CCAGCCGCAA GCGGCGCAAG CCGGTGCAGA AGACCGTGAA
60

GCC

63

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 65 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCACTAGTTG ATCACCATGA GCAGCGGCGC CAACATCACC TATGCCAGCC GCAAGCGCCG
60

26250 2045580

CAAGC

65

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCAGAGTCTG GGTTCAGAGC

20

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TCGAGTAGAT CACGCAATGG GCCCAGC

27

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TCGAGCTGGG CCCATTGCGT GATCTAC

20250519 10:54:02

27

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CGGTCGACTG GGCACCATGA ACAGCAGC
28

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCCAAGCTTA CGCGTGGTTC TCTGGAGGAA GCTGGTCTGG
40

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCCAAGCTTA CGCGTGGAAG TCTAGCTTGT GTTTGG
36

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCCAAGCTTA CGCGTGAAGC CGGAAACTG TCATGC
36

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCCAAGCTTA CGCGTGCAGT GGTCTCTGAG TGGCGATGAT GTAATCTGG
49

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCCAAGCTTA CGCGTGGTCT TTGAAGTCAA CCTCACC
37

(2) INFORMATION FOR SEQ ID NO:24:

20150510-051307

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAATTGTAAT ACGACTCACT ATAGGG
26

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CGCTCGAGAA CTAGTGGATC
20

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AGCTGCTTAA TTAATTAAGC A
21

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:

20250520 15:02:20

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AGCTTGCTTA ATTAATTAAG C

21

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCGTCGACTG ATGAGCAGCG GCGCCAACAT CACC

34

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CATTACTTAT CTAGAGCTCG

20

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs

20250510 15:13:27

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GATTTAGGTG AACTATAG

19

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GATAAGAATG CGGCCGCACG GATCCAGCAG CAACAGCAAA CAGAATTGG

49

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ATAAGAATGC GGCCGCAGCC CCCCCGACCG ATGTCAGC

38

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid

2025-10-20 15:13:37

- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ATAGTTTAGC GGCCGCCCCA CCGTACTCGT CAATTCC
37

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GCCGTCGACG CGGCCGCGAA GTCTAGCTTG TGTTTGG
37

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

ATAAGAATGC GGCCGCACCC TCAATGTTGT GTCGGG
36

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown

2045520-05199

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CGGGATCCTC GCGGCCGCAG AGAATTCAG GAATAGTGGC

40

0335402-051397